Printed: 24 November 2015

Chapter 5

Measurability woes

Section 5.1 *introduces three ways to overcome measurabilty difficulties that beset stochastic processes with uncountable index sets.*

Section 5.2 *explains how to replace supremum and infimum of uncountable families of random variables by weaker concepts—the essential supremum and infimum—that restore measurabilty at the cost of a few almost sure qualifiers.*

Section 5.3 *describes how to make negligible modifications to each member of an uncountable set of random variables, with each random variable being changed on its own negligible set, to produce a version with better sample path properties.*

Section 5.4 *describes ways of modifying the classical concept of convergence in distribution to accommodate the measurability difficulties caused by uncountable index sets.*

5.1 The difficulty

Versions::S:intro

Suppose $X = \{X_t : t \in T\}$ is a stochastic process, an indexed set of random variables all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If T is countable, the typical operations—sums, limits, products, suprema—do not take us outside the set of all ($\mathcal{F}$-measurable) random variables. If T is uncountable, questions of measurability become serious. For example, $\sup_{t\in T} X_t(\omega)$ need not be $\mathcal{F}$-measurable.

Probabilists have developed several strategies for dealing with the difficulties raised by uncountable T.

(i) Work with outer integrals and measurable cover functions. See Sections 5.2 and 5.4.

(ii) Replace each X_t by a new random variable $\widetilde{X}_t$, also defined on Ω, for which $\mathbb{P}\{\omega : X_t(\omega) \neq \widetilde{X}_t(\omega)\} = 0$ for each $t \in T$. Choose the new variables so that each sample path $\widetilde{X}(\omega, \cdot)$ is controlled (in some sense) by its behavior on a fixed, countable subset S of T. See Section 5.3.

(iii) Use the properties of analytic sets to establish measurability for quantities like $\sup_{t\in T} X_t(\omega)$ if $(\omega, t) \mapsto X(\omega, t)$ is product measurable and T can be identified with an analytic subset of a compact metric space. For the meaning of "analytic" see Dellacherie and Meyer (1978, Chapter III, no. 1 through 14). See also Dudley (1999, Chapter 5) and Pollard (1984, Appendix C).

In practice, for specific applications, I have found measurability problems easily handled by ad hoc approximation arguments using countable subsets of the index set. I have nothing to say in this book about method (iii), except that it lies mathematically deeper than the other two methods and that it becomes essential for a real understanding of stochastic calculus at the level of rigor of the Métivier (1982) book.

5.2 Essential supremum and infimum

Versions::S:essential

For a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$ write $\overline{\mathbb{M}}$ for the set of all $\mathcal{F}\backslash\mathcal{B}(\overline{\mathbb{R}})$-measurable maps from Ω into $\overline{\mathbb{R}}$.

Versions::ess.sup.inf

<1> **Theorem.** *For each $\mathbb{F} \subseteq \overline{\mathbb{M}}$ there exists a countable subset $\mathbb{F}_0$ of $\mathbb{F}$ such that $F := \sup \mathbb{F}_0$ belongs to $\overline{\mathbb{M}}$ and*

(i) $\mathbb{P}\{F \geq f\} = 1$ for each $f \in \mathbb{F}$

(ii) if G is another element of $\overline{\mathbb{M}}$ for which $\mathbb{P}\{G \geq f\} = 1$ for each $f \in \mathbb{F}$ then $\mathbb{P}\{G \geq F\} = 1$.

Proof The defining properties (i) and (ii) are unaffected by a monotone, one-to-one transformation such as arctan. Thus there is no loss of generality in assuming existence of a finite constant c for which $\sup_{f\in\mathbb{F}} |f| \leq c$.

Write $\mathcal{S}$ for the collection of all countable subsets of $\mathbb{F}$. For each S in $\mathcal{S}$ define $f_S(\omega) := \sup_{f\in S} f(\omega)$. The boundedness assumption ensures that the constant $\tau := \max\{\mathbb{P}f_S : S \in \mathcal{S}\}$ is also finite. Choose a sequence $\{S_n\}$ in $\mathcal{S}$ for which $\mathbb{P}f_{S_n} > \tau - n^{-1}$. The set $\mathbb{F}_0 := \cup_n S_n$ is countable. Define $F := f_{\mathbb{F}_0} := \sup_{f\in\mathbb{F}_0} f$. Then $\mathbb{P}F = \tau$.

For each g in $\mathbb{F}$ the set $\mathbb{F}_1 := \mathbb{F}_0 \cup \{g\}$ is countable and

$$\mathbb{P}F = \tau \geq \mathbb{P}f_{\mathbb{F}_1} = \mathbb{P}\max(g, F),$$

implying $F = \max(g, F)$ almost surely and $g \leq F$ almost surely.

For assertion (ii), note that $G \geq \sup_{f \in \mathbb{F}_0} f = F$ almost surely, because $\mathbb{F}_0$ is countable.

□

The set $\mathbb{F}_0$ is not unique, but property (ii) ensures that the function F is unique up to almost sure equivalence. It is called the ***essential supremum*** of $\mathbb{F}$ and is denoted by $\operatorname{ess\,sup} \mathbb{F}$. The function $-\operatorname{ess\,sup}\{-f : f \in \mathbb{F}\}$ is called the ***essential infimum*** of $\mathbb{F}$ and is denoted by $\operatorname{ess\,inf} \mathbb{F}$.

Remark. Here I am indulging in the usual abuse of referring to an equivalence class of functions as a function. Whenever I write $\operatorname{ess\,sup} \mathbb{F}$ (or $\operatorname{ess\,inf} \mathbb{F}$) I mean you to understand it is any function from the equivalence class, defined by some countable subset $\mathbb{F}_0$ of $\mathbb{F}$, as above.

The ***outer measure*** of a subset A of Ω is defined as

$$\mathbb{P}^* A := \inf\{\mathbb{P}B : A \subseteq B \in \mathcal{F}\}.$$

You should convince yourself that

$$\mathbb{P}^* A = \mathbb{P}A^* \qquad \text{where } A^* = \operatorname{ess\,inf}\{B \in \mathcal{F} : A \subseteq B\}.$$

The set A^* is called the ***measurable cover*** for A. It is unique up to almost sure equivalence: if $A \subseteq B \in \mathcal{F}$ then $A^* \leq B$ almost surely.

Similarly, if h is a (non measurable) function from Ω into $\overline{\mathbb{R}}$, its measurable cover is defined as

$$h^* := \operatorname{ess\,inf}\{f \in \overline{\mathbb{M}} : f(\omega) \geq h(\omega) \text{ for all } \omega \in \Omega\}.$$

Again the measurable cover is unique only up to almost sure equivalence. It is (almost surely) characterized by:

(i) $h^*(\omega) \geq h(\omega)$ for all $\omega \in \Omega$;

(ii) if g is a measurable function for which $g(\omega) \geq h(\omega)$ for all ω then $g \geq h^*$ almost surely.

Compare the definition of h^* with the definition of the outer integral:

$$\mathbb{P}^* h := \inf\{\mathbb{P}f : h \le f \in \overline{\mathcal{M}} \text{ and } \min\left(\mathbb{P}f^+, \mathbb{P}f^-\right) < \infty\}.$$

The last constraint on f ensures that $\mathbb{P}f$ is well defined, possibly taking the value $+\infty$ or $-\infty$. In general, $\mathbb{P}^*h$ need not equal $\mathbb{P}h^*$ (van der Vaart and Wellner, 1996, Problem 2 in Section 1.2). However if h is bounded—the only case needed in what follows—then there is equality. Proof?

The story is similar for lower integrals, $\mathbb{P}_* h = -\mathbb{P}^*(-h)$, and the lower analog of measurable covers, $h_* = -(-h)^*$.

I find that arguments involving measurable cover functions tend to be straightforward (but perhaps a little tedious) once I reduce the desired properties to assertions about measurable sets and measurable functions.

Versions::indic.mc <2> **Example.** For each real t show that $\{h^* > t\}$ is in the equivalence class of $\{h > t\}^*$. (Compare with Dudley, 1999, Lemma 3.2.6.)

By construction $h^* \ge h$ so $\{h^* > t\} \ge \{h > t\}$. The task reduces to showing: if $\{h > t\} \le B \in \mathcal{F}$ then $\{h^* > t\} \le B$ almost surely. Define a new measurable function

$$g(\omega) := h^*(\omega)\{\omega \in B\} + (t \wedge h^*(\omega))\{\omega \in B^c\}.$$

Clearly $g(\omega) \ge h(\omega)$ if $\omega \in B$. If $\omega \in B^c$ then $h(\omega) \le t$, so that $t \wedge h^*(\omega) \ge h(\omega)$. In short, $g(\omega) \ge h(\omega)$ for all ω, implying $g \ge h^*$ almost surely. That is, there exists a $\mathbb{P}$-negligible set $\mathcal{N}$ such that $g(\omega) \ge h^*(\omega)$ for all ω in $\mathcal{N}^c$.

If $\omega \in B^c \backslash \mathcal{N}$ then $t \wedge h^*(\omega) = g(\omega) \ge h^*(\omega)$. That is, $B^c \backslash \mathcal{N} \le \{h^* \le t\}$, which implies $\{h^* > t\} \le (B^c \backslash \mathcal{N})^c = B \cup \mathcal{N}$.

□

5.3 Separable versions

Versions::S:separable

The difficulties with an uncountable T can usually be handled by taking limits along a countable dense subset of T if the X process is separable, in the following sense.

Versions::sep <3> **Definition.** *Let S be a countable, dense subset S of a semimetric space T. Say that a process $X = \{X_t : t \in T\}$ is **S-separable** if for each ω in Ω and each t in T there exists a sequence $\{s_n : n \in \mathbb{N}\}$ (which might depend on ω) in S for which $s_n \to t$ and $X(\omega, s_n) \to X(\omega, t)$.*

Remark. Many authors allow a single $\mathbb{P}$-negligible set $\mathcal{N}$ such that the approximation property need only hold for ω in $\mathcal{N}^c$.

Not every process is separable. The classic example is the stochastic process $X(\omega, t) = 1$ if $t = \omega$, and zero otherwise, when $T = \Omega = [0, 1]$.

For many probabilistic purposes, two random variables Y and Z for which $\mathbb{P}\{\omega : Y(\omega) \neq Z(\omega)\} = 0$ are essentially the same. Indeed, many random variables are only defined up to some sort of almost sure equivalence. In isolation there is usually no good reason to prefer one choice from the almost sure equivalence class over another. However, for stochastic processes, whenever we need good behavior for the sample paths, $t \mapsto X(\omega, t)$ for each fixed ω, the selection from an equivalence class becomes much more important.

Versions::version <4> **Definition.** *Say that a stochastic process $\widetilde{X} = \{\widetilde{X}_t : t \in T\}$ is a version of $X = \{X_t : t \in T\}$ if they are both defined on the same probability space and $\mathbb{P}\{\omega : X_t(\omega) \neq \widetilde{X}_t(\omega)\} = 0$ for each t in T.*

> **Remark.** For each t there exists a $\mathbb{P}$-negligible set $\mathcal{N}_t$ for which $X_t(\omega) = \widetilde{X}_t(\omega)$ if $\omega \notin \mathcal{N}_t$. If T is countable, the set $\mathcal{N} := \cup_t \mathcal{N}_t$ is also $\mathbb{P}$-negligible; all differences between X and $\widetilde{X}$ appear only within a single negligible set $\mathcal{N}$. If T is uncountable, $\mathcal{N}$ need not be negligible.

For the traditional case where each X_t is an $\mathcal{F}$-measurable map into the real line it is usually necessary to allow $\widetilde{X}_t$ to take values in $\overline{\mathbb{R}} := [-\infty, +\infty]$. For X processes taking values in $\mathbb{R}^k$ the $\widetilde{X}$ process should be allowed to take values in $\overline{\mathbb{R}}^k$. Both cases are covered just by assuming that X takes values in some compact metric space, whose metric I denote by $\mathfrak{d}$ to avoid confusion with the semimetric d on T.

Remember that a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be ***complete*** if, for each $A \subseteq \Omega$, if $A \subseteq F \in \mathcal{F}$ and $\mathbb{P}F = 0$ then $A \in \mathcal{F}$.

Versions::sep.version <5> **Theorem.** *Suppose T is a separable semimetric space and each X_t takes values in a compact metric space $(E, \mathfrak{d})$. If the underlying probability space is complete then there exists a version of X that is separable. That is, there exists a version $\widetilde{X}$ of X and a countable dense subset S of T such that every sample path $\widetilde{X}(\omega, \cdot)$ is S-separable in the sense of Definition <3>.*

Proof There exists a countable collection $\mathcal{U}$ of open subsets of T such that $G = \bigcup\{U \in \mathcal{U} : U \subseteq G\}$ for every open set G. In particular, for each t in T there exists a decreasing sequence of sets $U_n(t)$ from $\mathcal{U}$ for which $t \in \cap_n U_n(t)$ and $\text{diam}(U_n(t)) \to 0$ as $n \to \infty$.

Let A be any countable dense subset of E. For each t in T and $\alpha \in A$ write $f_{t,\alpha}(\omega)$ for $\mathfrak{d}\,(X_t(\omega), \alpha)$. For each $U \in \mathcal{U}$ and $\alpha \in A$, define

$$Y_{U,\alpha}(\omega) = \operatorname*{ess\,inf}_{t \in U} f_{t,\alpha}(\omega) = \inf_{s \in S(U,\alpha)} f_{s,\alpha}(\omega),$$

where $S(U,\alpha)$ is a countable subset of U. Remember that for each $t \in U$ there exists a $\mathbb{P}$-negligible set $\mathcal{N}_{t,U,\alpha}$ for which

$$f_{t,\alpha}(\omega) \geq Y_{U,\alpha}(\omega) \qquad \text{for all } \omega \notin \mathcal{N}_{t,U,\alpha}$$

Define $S := \bigcup\{S(U,\alpha) : U \in \mathcal{U},\ \alpha \in A\}$ and $\widetilde{X}(\omega,s) = X(\omega,s)$ for all ω and all $s \in S$.

For each fixed t in $T\backslash S$, the set

$$\mathcal{N}_t := \bigcup\{\mathcal{N}_{t,U,\alpha} : t \in U \in \mathcal{U},\ \alpha \in A\}.$$

is $\mathbb{P}$-negligible. Completeness of the probability space ensures that we can change $X(\omega,t)$ for $\omega \in \mathcal{N}_t$ anyway we like without upsetting the measurability.

More precisely, for $\omega \notin \mathcal{N}_t$ define $\widetilde{X}(\omega,t) = X(\omega,t)$ and for $\omega \in \mathcal{N}_t$ define $\widetilde{X}(\omega,t)$ to equal some arbitrarily chosen point in the compact set

$$\bigcap\nolimits_{n\in\mathbb{N}} \overline{\{X(\omega,s) : s \in S \cap U_n(t)\}}$$

Remark. If E were not compact the intersection of a decreasing sequence of closed sets might be empty.

To see that the $\widetilde{X}$ process is S-separable, once again consider a t in $T\backslash S$. If $\omega \in \mathcal{N}_t$, by definition of the closure there exist points $s_n \in S \cap U_n(t)$ for which

$$\mathfrak{d}(\widetilde{X}(\omega,t), X(\omega,s_n) < n^{-1} \qquad \text{for } n \in \mathbb{N},$$

so that $s_n \to t$ and $\widetilde{X}(\omega,s_n) = X(\omega,s_n) \to \widetilde{X}(\omega,t)$ as $n \to \infty$. For $\omega \notin \mathcal{N}_t$ there exists a sequence $\{\alpha_n\}$ in E for which $\alpha_n \to X(\omega,t)$. For each n there exists an s_n in $S(U_n(t),\alpha_n)$ for which

$$n^{-1} + \mathfrak{d}(X(\omega,t),\alpha_n) \geq n^{-1} + Y_{U_n(t),\alpha_n}(\omega) \geq \mathfrak{d}(X(\omega,s_n),\alpha_n),$$

ensuring that $s_n \to t$ and $\widetilde{X}(\omega,s_n) = X(\omega,s_n) \to X(\omega,t) = \widetilde{X}(\omega,t)$.

□

Recall that the oscillation on a subset S of T for a function f from T to the real line is defined by

$$\text{osc}(\delta,f,S) = \sup\{|f(t) - f(t')| : t,t' \in S,\ d(t,t') < \delta\}.$$

If f takes values in a metric space $(E, \mathfrak{d})$, the quantity $|f(t) - f(t')|$ should be replaced by $\mathfrak{d}(f(t), f(t'))$. If $\operatorname{osc}(\delta, f, T) \to 0$ as $\delta \to 0$ then f is uniformly continuous on T.

For a stochastic process $\{X_t : t \in T\}$ the oscillation is defined for each sample path $X_\omega := X(\omega, \cdot)$. Unfortunately, the map $\omega \mapsto \operatorname{osc}(\delta, X_\omega, t)$ might not be measurable if T is uncountable. The difficulty disappears if X is S-separable for some countable S, because

$$\operatorname{osc}(\delta, X_\omega, T) = \operatorname{osc}(\delta, X_\omega, S) \qquad \text{for each } \omega.$$

Indeed, if $d(t, t') < \delta$ then there exist sequences $\{s_n\}$ and $\{s'_n\}$ in S for which $d(s_n, t) \to 0$ and $d(s'_n, t') \to 0$ and $X_\omega(s_n) \to X_\omega(t)$ and $X_\omega(s'_n) \to X_\omega(t')$. For large enough n, we have $d(s_n, s'_n) < \delta$ and

$$\mathfrak{d}(X_\omega(s_n), X_\omega(s'_n)) \to \mathfrak{d}(X_\omega(t), X_\omega(t')).$$

Versions::cts.paths <6> **Example.** Suppose $\{X_t : t \in T\}$ is an S-separable takes values in a metric space $(E, \mathfrak{d})$. Suppose also that for each $\eta > 0$ and $\epsilon > 0$ there exists a $\delta > 0$ for which

\E@ X.osc <7>
$$\mathbb{P}\{\omega \in \Omega : \operatorname{osc}(\delta, X_\omega, S) > \eta\} < \epsilon.$$

By separability the same inequality holds with S replaced by T.

Let δ_k denote the value of δ corresponding to $\eta = k^{-1}$ and $\epsilon = 2^{-k}$ in <7>. Then $\sum_{k \in \mathbb{N}} \mathbb{P}\{\operatorname{osc}(\delta_k, \widetilde{X}_\omega, T) > k^{-1}\} < \infty$. By Borel-Cantelli there exists a single $\mathbb{P}$-negligible set $\mathcal{N}$ such that $\operatorname{osc}(\delta_k, \widetilde{X}_\omega, T) \to 0$ as $n \to \infty$ if $\omega \notin \mathcal{N}$. Amost all sample paths are uniformly continuous as functions on T.

□

If the X process in Example <6> were not separable, but if it it had an S-separable version $\widetilde{X}$ then almost all sample paths of $\widetilde{X}$ would be uniformly continuous as functions on T. Of course to guarantee existence of the separable version we might need the range metric space E to be compact. For example, if X took values in $\mathbb{R}$ the $\widetilde{X}$ process might take values in $\overline{\mathbb{R}}$. However, a proof of a condition like <7> typically requires T to be totally bounded, that is, for each $\delta > 0$ the set T can be covered by a finite union of δ-balls. In that case $\widetilde{X}$ will take values in $\mathbb{R}$, because total boundedness and $\operatorname{osc}(\delta, X_\omega, S) < \infty$ implies $\sup_{s \in S} |X(\omega, s)| < \infty$. Total boundedness also implies that the completion $\overline{T}$ of T is compact (Dudley, 2003, Theorem 2.5.1). Every sample path $\widetilde{X}_\omega$ then has a unique extension to a uniformly continuous, real-valued function on $\overline{T}$.

5.4 Convergence in distribution

Versions::S:cid

For this section suppose T is a totally bounded, (uncountable) semimetric space. Write $\text{UC}(T,d)$ for the set of all uniformly continuous real functions on T. Total boundedness of T ensures that each function in $\text{UC}(T,d)$ also belongs to the vector space $\ell^\infty(T)$ of all bounded, real-valued functions on T. Equip $\ell^\infty(T)$ (and $\text{UC}(T,d)$) with the uniform norm $\|\cdot\|_T$ and the ***cylinder sigma-field*** $\mathcal{C}_T$, the smallest sigma-field that makes each coordinate map measurable.

The metric space $\ell^\infty(T)$ is complete under the uniform norm, but it is not separable if T is uncountable. The Borel sigma-field on $\ell^\infty(T)$ is larger than $\mathcal{C}_T$. The space $\text{UC}(T,d)$ is a closed subset of $\ell^\infty(T)$. As a metric space in its own right, $\text{UC}(T,d)$ is complete and separable and its Borel sigma-field $\mathcal{B}_{\text{UC}}$ coincides with its cylinder sigma-field, $\mathcal{C}_T$ (Problem [5] and [6]).

Write $X^{(\infty)}$ for the S-separable process described in Example <6>. Each sample path of $X^{(\infty)}$ belongs to $\text{UC}(T,d)$ and for each $\epsilon>0$ and $\eta>0$ there exists a $\delta>0$ for which

\E@ Xinfty.osc <8>
$$\mathbb{P}\{\text{OSC}(\delta, X^{(\infty)}, T) > \eta\} < \epsilon.$$

Measurabilty of each $X_t^{(\infty)}$ ensures that $X^{(\infty)}$ is $\mathcal{F}\backslash\mathcal{C}_T$-measurable as a random element of $\ell^\infty(T)$ or of $\text{UC}(T,d)$. As a random element of $\text{UC}(T,d)$ it is also $\mathcal{F}\backslash\mathcal{B}_{\text{UC}}$-measurable.

Now suppose $X^{(n)} = \{X_t^{(n)} : t \in T\}$ is a sequence of stochastic processes, all living on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and sharing the same index set. Assume that $\sup_{t\in T}|X^{(n)}(\omega,t)| < \infty$ for each ω and n, so that each $X^{(n)}$ is also an $\mathcal{F}\backslash\mathcal{C}_T$-measurable random element of $\ell^\infty(T)$.

Suppose also that $X^{(n)} \rightsquigarrow_{fidi} X^{(\infty)}$, meaning that the finite dimensional distributions of the $X^{(n)}$'s converge to the corresponding finite dimensional distributions of $X^{(\infty)}$: for each finite subset $\{t_1,\dots,t_k\}$ of T,

$$(X^{(n)}(t_1),\dots,X^{(n)}(t_k)) \rightsquigarrow (X^{(\infty)}(t_1),\dots,X^{(\infty)}(t_k))\,,$$

where $\rightsquigarrow$ denotes "convergence in distribution" (Pollard, 2001, Chapter 7).

What more than finite dimensional convergence is needed to establish something like $\mathbb{P}f(X^{(n)}) \to \mathbb{P}f(X^{(\infty)})$ for a broad collection of (bounded) functionals f, which assign a real number to each sample path?

The main difficulty in what follows is: $\text{OSC}(\delta, x, T)$ is a uniformly continuous function of x on $\ell^\infty(T)$, and hence it is measurable with respect to the Borel sigma-field $\mathcal{B}(\ell^\infty(T))$, which can be much larger than $\mathcal{C}_T$ when T is uncountable. If $X^{(n)}$ were not $\mathcal{F}\backslash\mathcal{B}(\ell^\infty(T))$-measurable then $\text{OSC}(\delta, X^{(n)}, T)$

might not be an $\mathcal{F}$-measurable random variable, which would complicate the theory for convergence in distribution.

To explain the consequences of this measurabilty obstacle, I separate the discussion into three related possibilities. If you get bored by the repetition you might skip straight to the last case, even though that might leave you wondering why such a strange definition has become established in the literature.

5.4.1 Processes with uniformly continuous paths

Versions::cid.ucts

Suppose each $X^{(n)}$ has sample paths in $\text{UC}(T,d)$. That is, each $X^{(n)}$ (and $X^{(\infty)}$) defines an $\mathcal{F}\backslash\mathcal{C}_T$-measurable random element of $\text{UC}(T,d)$. In this case there is no measurability difficulty with the oscillation functions: it falls into the classical case where the convergence in distribution $X^{(n)} \rightsquigarrow X^{(\infty)}$ means $\mathbb{P}f(X^{(n)}) \to \mathbb{P}f(X^{(\infty)})$ for each bounded Lipschitz real function on $\text{UC}(T,d)$. The convergence extends to other bounded functionals by means of the usual semi-continuity and almost sure continuity arguments (Pollard, 2001, Section 7.1).

The necessary and sufficient condition for the convergence $X^{(n)} \rightsquigarrow X^{(\infty)}$ is $X^{(n)} \rightsquigarrow_{fidi} X^{(\infty)}$ plus an equicontinuity condition: for each $\epsilon > 0$ and $\eta > 0$ there exists a $\delta > 0$ for which

\E@ ucts.equicty

$$\limsup_{n\to\infty} \mathbb{P}\{\text{OSC}(\delta, X^{(n)}, T) > \eta\} < \epsilon \tag{9}$$

Let me remind you of one way to establish sufficiency.

For each $\delta > 0$ Problem [4] shows there exists a finite subset T_δ of T and a map $\mathbb{A}_\delta : \ell^\infty(T) \to \text{UC}(T,d)$ for which

(i) $\mathbb{A}_\delta(x)$ is a Lipschitz function of $x(T_\delta) := (x(t) : t \in T_\delta)$

(ii) $\|x - \mathbb{A}_\delta x\|_T \le \text{OSC}(2\delta, x, T)$.

If f is a bounded Lipschitz real function on $\ell^\infty(T)$ (or even just on $\text{UC}(T,d)$) then there exists a finite constant L for which

$$|f(x) - f(y)| \le L\,(1 \wedge \|x-y\|_T)\,.$$

From (i), the real number $f(\mathbb{A}_d x)$ is a bounded Lipschitz function of $x(T_\delta)$, so that $\mathbb{P}f(\mathbb{A}_\delta X^{(n)}) - \mathbb{P}f(\mathbb{A}_\delta X^{(\infty)})$ tends to zero as $n \to \infty$. Also

$$|f(x) - f(\mathbb{A}_\delta x)| \le L\,(1 \wedge \text{OSC}(2\delta,x,T)) \le L\,(\eta + \{\text{OSC}(2\delta,x,T) > \eta\})\,.$$

A simple triangle inequality argument gives

$$\begin{aligned}|\mathbb{P}f(X^{(n)}) - \mathbb{P}f(X^{(\infty)})| \le \mathbb{P}|f(X^{(n)}) - f(\mathbb{A}_\delta X^{(n)})| \\ + |\mathbb{P}f(\mathbb{A}_\delta X^{(n)}) - \mathbb{P}f(\mathbb{A}_\delta X^{(\infty)})| \\ + \mathbb{P}|f(\mathbb{A}_\delta X^{(\infty)})) - f(X^{(\infty)})|.\end{aligned}$$

By <9> (with δ replaced by 2δ) and its analog for $X^{(\infty)}$, the first and third terms on the right-hand side are each eventually less than $L(\eta+\epsilon)$ and the middle term tends to zero. In short, $\mathbb{P}f(X^{(n)}) \to \mathbb{P}f(X^{(\infty)})$ as $n \to \infty$.

5.4.2 Separable versions

Versions::cid.separable

Write $\mathcal{S}$ for the collection of all countable subsets of T. If $X^{(n)}$ is not Borel measurable there is still hope of proving that for each $\epsilon > 0$ and $\eta > 0$ there exists a $\delta > 0$ for which

\E@ ucts.equicty2

<10> $$\limsup_{n\to\infty} \mathbb{P}\{\text{OSC}(\delta, X^{(n)}, S) > \eta\} < \epsilon \qquad \text{for each } S \in \mathcal{S}.$$

This inequality also fits with the philosophy that, without extra sample path properties, we do not really understand much about a stochastic process except for almost sure facts about behavior on finite or countable subsets of its index set. However, if we do know that each $X^{(n)}$ is S-separable for a fixed S in $\mathcal{S}$ then much of the argument from the previous section can be rescued. For example,

$$\text{OSC}(\delta, X_\omega^{(n)}, T) = \text{OSC}(\delta, X_\omega^{(n)}, S) \qquad \text{for all } \omega, \text{ all } n, \text{ all } \delta > 0.$$

Thus each $\text{OSC}(\delta, X^{(n)}, T)$ is $\mathcal{F}$-measurable. However, the random variable $f(X^{(n)})$, for a bounded Lipschitz f, need not be $\mathcal{F}$-measurable.

We do still have an upper bound for the measurable cover of $f(X^{(n)})$:

$$f(X^{(n)})^* \le f(\mathbb{A}_\delta X^{(n)}) + L\left(\eta + \{\text{OSC}(2\delta, X^{(n)}, T) > \eta\}\right).$$

so that

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}^* f(X^{(n)}) &= \limsup_{n\to\infty} \mathbb{P}f(X^{(n)})^* \\ &\le \mathbb{P}f(\mathbb{A}_\delta X^{(\infty)}) + L\,(\eta+\epsilon) \\ &\le \mathbb{P}f(X^{(\infty)}) + 2L\,(\eta+\epsilon)\end{aligned}$$

Thus $\limsup_{n\to\infty} \mathbb{P}^* f(X^{(n)}) \le \mathbb{P}f(X^{(\infty)})$ for all bounded, Lipschitz f. Replace f by $-f$ to deduce that the last inequality is equivalent to the assertion $\liminf_{n\to\infty} \mathbb{P}_* f(X^{(n)}) \ge \mathbb{P}f(X^{(\infty)})$. At least in the sense of the following Definition, we can still deduce convergence in distribution (for separable process) from finite dimensional convergence plus <10>.

Versions::cid.outer

<11> **Definition.** *If X is a Borel measurable random element of $\ell^\infty(T)$ and $\{X^{(n)}\}$ is a sequence of $\mathcal{C}_T$-measurable random elements of $\ell^\infty(T)$, define the convergence in distribution $X^{(n)} \rightsquigarrow X^{(\infty)}$ to mean $\limsup_{n\to\infty} \mathbb{P}^* f(X^{(n)}) \le \mathbb{P} f(X)$ for all bounded, Lipschitz functions f from $\ell^\infty(T)$ to $\mathbb{R}$.*

5.4.3 Outer expectations

Versions::cid.outer.sub

For some stochastic processes, such as the classical empirical processes, many authors have objected to the implicit modification that replaces a process $X^{(n)}$ by a separable version, on the grounds that it radically changes some of the properties of the process. For example, if $X_n(t) = \sqrt{n}(P_n - P)f_t$ then a separable version need not retain the interpretation of an integral with respect to a random signed measure.

For that situation it seems inevitable that assumption <10> be replaced by its outer measure analog: for each $\epsilon > 0$ and $\eta > 0$ there exists a $\delta > 0$ for which

\E@ ucts.equicty.outer

$$\limsup_{n\to\infty} \mathbb{P}^*\{\text{OSC}(\delta, X^{(n)}, T) > \eta\} < \epsilon. \qquad <12>$$

Such an assumption opposes the philosophy of knowing only about finite or countable subsets of T. Nevertheless, it still leads to full convergence in distribution. I leave the details of the argument to you—it involves only an addition of a few $*$'s to the argument in the previous subsection. Moreover, as you will see in later chapters, it is not so outrageous to work with $\text{OSC}(2\delta, x, T)^*$ because we usually have to control the oscillation by means of an upper bound that involves only measurable functions and fairly routine passages from bounds for countable index subsets to uncountable index sets.

5.5 Problems

Versions::S:problems

Versions::P:pack.cty

[1] (from Chapter 4) Suppose $\|X_s - X_t\|_{\Psi_\alpha} \le d(s,t)$ for all $s, t \in T$. Suppose also that

$$\int_0^D \Psi_\alpha^{-1}(\text{PACK}(r, T, d))\, dr < \infty.$$

Prove that X has a version with uniformly continuous sample paths.

Versions::P:bdd.mble

[2] Suppose h is a bounded real function on Ω and $(\Omega, \mathcal{F}, \mathbb{P})$ is complete. Show that h is $\mathcal{F}$-measurable if and only if $\mathbb{P}\left(h^* + (-h)^*\right) = 0$.

Versions::P:complete.T

[3] Suppose T is a semimetric space and $\overline{T}$ is its completion. Show that every function x in $\text{UC}(T,d)$ has a unique extension to a function in $\text{UC}(\overline{T},d)$. (Here d denotes both the semimetric on T and its unique extension to a semimetric on $\overline{T}$.)

Versions::P:ucts.approx

[4] Suppose T is a totally bounded semimetric space. Given $\delta > 0$, let $T_\delta = \{t_1, \dots, t_N\}$ be a maximal δ-separated subset of T, so that $d(s, T_\delta) := \min\{d(s,t_i) : 1 \le i \le N\} \le \delta$ for every s in T.

(i) Define $f_i(t) := (1 - d(t,t_i)/(2\delta))^+$ and $F(t) := \sum_{i \le N} f_i(t)$. Show that $F(t) \ge 1/2$ for all t and $g_i(t) := f_i(t)/F(t)$ is uniformly continuous.

(ii) Define a map $A_\delta : \mathbb{R}^N \to \text{UC}(T,d)$ by $A_\delta(\alpha) := \sum_{i \le N} \alpha_i g_i(s)$. Also define $\mathbb{A}_\delta : \text{UC}(T,d) \to \text{UC}(T,d)$ by $\mathbb{A}_\delta(x) := A_\delta(x(T_\delta))$, the function that takes the value $\sum_{i \le N} x(t_i) g_i(t)$ at t. Show that $\|x - \mathbb{A}_\delta(x)\|_T \le \text{OSC}(2\delta, x, T)$. Hint: If $g_i(t) > 0$ then $d(t, t_i) < 2\delta$.

(iii) For all x, y in $\text{UC}(T,d)$, show that

$$\|\mathbb{A}_\delta(x) - \mathbb{A}_\delta(y)\|_T \le \max\nolimits_{i \le N} |x(t_i) - y(t_i)| \le \|x - y\|_T .$$

Versions::P:sep.ucts

[5] Suppose T is a totally bounded semimetric space. Show that $\text{UC}(T)$ is separable under its uniform metric. Hint: For a sequence $\{\delta_k\}$ tending to zero, consider the countable set of functions of the form $A_{\delta_k}(\alpha)$ (as in [4]) with all coordinates of α rational.

sions::P:ucts.sigma.fields

[6] Show that the Borel sigma-field $\mathcal{B}(T)$ generated by the norm $\|\cdot\|_T$ on $\text{UC}(T,d)$ is the same as the cylinder sigma-field $\mathcal{C}_T$. Hint: First show that $\|x\|_T$ equals a supremum taken over a fixed countable dense subset of T. Then show that each coordinate projection, $x \mapsto x(t)$ for fixed $t \in T$, is continuous.

Versions::P:f.cylinder

[7] Suppose <10> holds and $X^{(n)} \rightsquigarrow_{fidi} X^{(\infty)}$, where $X^{(\infty)}$ has sample paths in $\text{UC}(T,d)$, as in Section 5.4. Suppose also that f is a $\mathcal{C}_T$-measurable real-valued function on $\ell^\infty(T)$ for which $|f(x) - f(y)| \le L\,(1 \wedge \|x - y\|_T)$ for all $x, y \in \ell^\infty(T)$. For each subset A of T define $\|x\|_A := \sup_{t \in A} |x(t)|$ and write $\mathcal{C}_A$ for the smallest sigma-field that makes all the coordinate maps $x \mapsto x(t)$, for $t \in A$, measurable.

(i) Show that $\mathcal{C}_T = \cup_{S \in \mathcal{S}} \mathcal{C}_S$, where $\mathcal{S}$ is the collection of all countable subsets of T. Deduce that f is $\mathcal{C}_S$-measurable for some S in $\mathcal{S}$. The following parts all refer to that particular S.

(ii) Show that the Borel sigma-field on $\ell^\infty(S)$ is equal to sigma-field $\mathcal{C}_S$ generated by all the coordinate maps. (If you feel that the use of $\mathcal{C}_S$ as the name for a sigma-field on $\ell^\infty(S)$ and a different, but related, sigma-field on $\ell^\infty(T)$, you should invent some more precise notation.)

(iii) Write x_S for the restriction of a function x in $\ell^\infty(T)$ to the subset S. Note that $x_S \in \ell^\infty(S)$. Show that $f(x) = f(y)$ iff $x_S = y_S$.

(iv) Show that there exists a $\mathcal{C}_S$-measurable map f_0 from $\ell^\infty(S)$ into $\mathbb{R}$ for which $f(x) = f_0(x_S)$ for all $x \in \ell^\infty(T)$ and

$$|f_0(x_0) - f_0(y_0)| \le L\,(1 \wedge \|x_0 - y_0\|_S)$$

for all $x_0, y_0 \in \ell^\infty(S)$.

(v) Show that $\mathbb{P}f(X^{(n)}) = \mathbb{P}f_0(X_S^{(n)}) \to \mathbb{P}f_0(X_S^{(\infty)}) = \mathbb{P}f(X^{(\infty)})$.

(vi) Write a long essay discussing whether the $\mathcal{C}_T$-measurability of functionals on $\ell^\infty(T)$ provides a worthwhile solution to the problems caused by uncountable index sets.

5.6 Notes

Versions::S:Notes

For Section 5.2 I am mostly following Dudley (1999, Section 3.2).

Doob (1953, Section 1.2) described the virtues of working with separable versions. I do not know whether that was the original source for the idea, which is invariably attributed to Doob. I learned about the extension to processes taking values in compact metric spaces from Meyer (1966, Chapter IV.2). For clear accounts of the construction of separable versions see Neveu (1965, Section III.4) and Gihman and Skorohod (1974, Section III.2).

Separable versions can also be constructed in great generality by means of liftings, maps from the set of equivalance classes $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ into the set of bounded measurable functions that preserve the interesting operations (linearity, products, maximima). See Jacobs (1978, Chapter XV) for the construction and application of liftings, including existence of separable versions of stochastic processes taking values in completely regular Hausdorff spaces. I found the treatment by Jacobs more accessible than the traditional refernce for lifting, the book by Ionescu Tulcea and Ionescu Tulcea (1969).

In the setting of the classical Donsker theorem, Talagrand (1987) also used liftings to break the empirical process into a sum of two parts: one a separable version of the process and the other a process that is mostly zero. The second process captures all the measurabilty subtleties caused by an uncountable index set. Talagrand (1988) explained why some care is needed in the choice of lifting to avoid some undesirable empirical process properties.

The approach to convergence in distribution in function spaces is almost part of the probability folklore: the conditions for convergence in distribution are just the conditions for existence (of versions with sample paths in that function space) somehow made uniform over the sequence of processes. I learned to appreciate this point of view by reading some unpublished lecture notes by Peter Gänssler. A revised version of the notes appeared in Chapter 7 of the book by Gänssler and Stute (1977). Compare with Gihman and Skorohod (1974, Chapter III).

Definition <11> represents the subtle conclusion of a long line of attempts to handle the measurability problem gracefully. According to Dudley (1999, Notes to section 3.1) and van der Vaart and Wellner (1996, page 96), this particular combination of ideas is due to Hoffmann-Jørgensen, from some time in the mid 1980's. I believe the manuscript by Hoffmann-Jørgensen (1984) is the source. The success of Dudley's almost uniform representation theorem (Dudley, 1999, Section 3.5) convinced me that the Hoffmann-Jørgensen approach was the right one. See also (Pollard, 1990, Chapter 9) and Kim and Pollard (1990, Section 2).

References

DellacherieMeyer78bookA Dellacherie, C. and P. A. Meyer (1978). *Probabilities and Potential*. Amsterdam: North-Holland. (First of three volumes).

Doob53book Doob, J. L. (1953). *Stochastic Processes*. New York: Wiley.

Dudley1999UCLT Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.

Dudley2003RAP Dudley, R. M. (2003). *Real Analysis and Probability* (2nd ed.). Cambridge studies in advanced mathematics. Cambridge University Press.

GaensslerStute1977book Gänssler, P. and W. Stute (1977). *Wahrscheinlichkeitstheorie*. Springer-Verlag.

GihmanSkorohod:vol1 Gihman, I. I. and A. V. Skorohod (1974). *The Theory of Stochastic Processes*, Volume I. Springer.

HoffmannJorgensen84polish Hoffmann-Jørgensen, J. (1984). Stochastic Processes on Polish Spaces. Unpublished manuscript, Aarhus University, Denmark.

TulceaTulcea69book Ionescu Tulcea, A. I. and C. T. Ionescu Tulcea (1969). *Topics in the Theory of Lifting*. Springer-Verlag.

Jacobs1978

Jacobs, K. (1978). *Measure and Integral.* Academic Press.

KimPollard90cuberoot

Kim, J. and D. Pollard (1990). Cube root asymptotics. *Annals of Statistics 18*, 191–219.

Metivier82book

Métivier, M. (1982). *Semimartingales: A Course on Stochastic Processes.* Berlin: De Gruyter.

Meyer1966PP

Meyer, P. A. (1966). *Probability and Potentials.* Blaisdell.

Neveu65book

Neveu, J. (1965). *Mathematical Foundations of the Calculus of Probability.* San Francisco: Holden-Day.

Pollard84book

Pollard, D. (1984). *Convergence of Stochastic Processes.* New York: Springer.

Pollard90Iowa

Pollard, D. (1990). *Empirical Processes: Theory and Applications*, Volume 2 of *NSF-CBMS Regional Conference Series in Probability and Statistics.* Hayward, CA: Institute of Mathematical Statistics.

PollardUGMTP

Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability.* Cambridge University Press.

Talagrand1987mble

Talagrand, M. (1987). Measurability problems for empirical processes. *Annals of Probability 15*(1), 204–212.

Talagrand1988liftings

Talagrand, M. (1988). On liftings and the regularization of stochastic processes. *Probability theory and related fields 78*(1), 127–134.

vaartwellner96book

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Process: With Applications to Statistics.* Springer-Verlag.